

The effect of school attendance and school dropout on incident HIV and HSV-2 among young women in rural South Africa enrolled in HPTN 068

Marie C.D. Stoner^a, Audrey Pettifor^{a,b}, Jessie K. Edwards^a, Allison E. Aiello^a, Carolyn T. Halpern^c, Aimée Julien^{a,b}, Amanda Selin^a, Rhian Twine^b, James P. Hughes^{e,f}, Jing Wang^f, Yaw Agyei^g, F. Xavier Gomez-Olive^{b,h}, Ryan G. Wagner^{b,d}, Catherine MacPhail^{b,i,j} and Kathleen Kahn^{b,d,h}

Objective: To estimate the association between school attendance, school dropout, and risk of incident HIV and herpes simplex virus type 2 (HSV-2) infection among young women.

Design: We used longitudinal data from a randomized controlled trial in rural Mpumalanga province, South Africa, to assess the association between school days attended, school dropout, and incident HIV and HSV-2 in young women aged 13–23 years.

Methods: We examined inverse probability of exposure weighted survival curves and used them to calculate 1.5, 2.5, and 3.5-year risk differences and risk ratios for the effect of school attendance on incident HIV and HSV-2. A marginal structural Cox model was used to estimate hazard ratios for the effect of school attendance and school dropout on incident infection.

Results: Risk of infection increased over time as young women aged, and was higher in young women with low school attendance (<80% school days) compared with high (≥80% school days). Young women with low attendance were more likely to acquire HIV [hazard ratio (HR): 2.97; 95% confidence interval (CI): 1.62, 5.45] and HSV-2 (HR: 2.47; 95% CI: 1.46, 4.17) over the follow-up period than young women with high attendance. Similarly, young women who dropped out of school had a higher weighted hazard of both HIV (HR 3.25 95% CI: 1.67, 6.32) and HSV-2 (HR 2.70; 95% CI 1.59, 4.59).

Conclusion: Young women who attend more school days and stay in school have a lower risk of incident HIV and HSV-2 infection. Interventions to increase frequency of school attendance and prevent dropout should be promoted to reduce risk of infection.

Copyright © 2017 Wolters Kluwer Health, Inc. All rights reserved.

AIDS 2017, **31**:2127–2134

Keywords: education, HIV, herpes simplex virus type 2, South Africa, young women

^aDepartment of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, USA, ^bMRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), Faculty of Health Sciences, School of Public Health, University of the Witwatersrand, Johannesburg, South Africa, ^cDepartment of Maternal and Child Health, University of North Carolina, Chapel Hill, North Carolina, USA, ^dEpidemiology and Global Health Unit, Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden, ^eDepartment of Biostatistics, University of Washington, ^fFred Hutchinson Cancer Research Center, Seattle, Washington, ^gDepartment of Pathology, School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA, ^hINDEPTH Network, Accra, Ghana, West Africa, ⁱSchool of Health and Society, University of Wollongong, New South Wales, Australia, and ^jWits Reproductive Health and HIV Research Institute, University of the Witwatersrand, Johannesburg, South Africa. Correspondence to Marie C.D. Stoner, 135 Dauer Drive, 2101 McGavran-Greenberg Hall, Chapel Hill, NC 27599-7435, USA.

E-mail: stonerm@email.unc.edu

Received: 10 April 2017; revised: 20 June 2017; accepted: 26 June 2017.

Introduction

Young South African women are disproportionately at risk of HIV and herpes simplex virus type 2 (HSV-2). In rural South Africa, the prevalence of HIV is 16% among young women aged 15 to 24 compared to 3% among young men [1]. Estimates of the prevalence of HSV-2 in this population are higher than HIV and show a similar differential by sex; 29% among young women and 10% among young men aged 15–26 [2]. Increased education may prevent HIV and HSV-2. Higher levels of education have been associated with a reduced prevalence [3–5] and incidence [6,7] of HIV in several countries in sub-Saharan Africa with the largest decreases occurring in younger, more educated women. In Botswana, each additional year of secondary schooling resulted in an 8.1% reduction in the cumulative risk of HIV infection overall and 11.6% in women [6]. However, all prior studies linking level of education to HIV and HSV-2 have been among adults who have completed their education. Few studies directly examine the relationship between school attendance and HIV infection among adolescents of school age despite current HIV prevention policies to keep girls in school [8].

Highest level of education completed, or educational attainment, is a measure of cumulative exposure to education over one's lifetime and encompasses life skills and knowledge gained as well as other factors that occur across the life course that are linked to education such as increased socioeconomic status (SES) [9]. Interventions to keep girls in school will not only increase educational attainment, but will more immediately influence factors related to being in a school environment that also affect sexually transmitted infections (STI) acquisition, such as sexual networks and having less time to engage in sexual activity [9]. The impact of school attendance among young women is often not studied as an independent factor from educational attainment. Distinguishing these parameters may provide additional insight into how structured time in school versus cumulative gains in knowledge influence risk of infection.

Although there is evidence that attending school reduces prevalent HIV and HSV-2 infection among young women, there is neither longitudinal evidence nor incident analysis to support this relationship [10–14]. Our study uses longitudinal data from a randomized controlled trial of young women in South Africa to assess the effect of school attendance and school dropout on incident HIV and HSV-2. We hypothesize that young women who stay in school and attend more days of school will have a lower risk of incident HIV and HSV-2 infection compared with young women who attend less school days or who dropout.

Methods

We used longitudinal data from the HIV Prevention Trials Network (HPTN) 068 study, a phase III randomized trial to determine whether providing cash transfers, conditional on school attendance, reduces young women's risk of HIV acquisition [15,16]. The study enrolled 2533 young women aged 13–20 years who were attending high school grades 8–11 in the Bushbuckridge subdistrict in Mpumalanga province, South Africa. The study setting is rural with poor infrastructure and is characterized by high levels of poverty, unemployment, and migration for work [15,17,18]. In 2007, the percentage in each village under the poverty line ranged from 15 to 81% [19]. Young women have almost universal access to primary education but the quality of education is poor and progress is often delayed [18]. Almost all villages in the study area have a high school. The most common reason for school dropout is being pregnant or having a child and the most common reason for low attendance is being sick or disabled followed by having to help at home [Stoner *et al.*, forthcoming].

The original HPTN 068 study excluded young women who were pregnant or married at baseline or had no parent/guardian in the household. Young women who dropped out of school during the study period were not removed from the study. Our analysis further excluded young women without a follow-up HIV test after baseline and those with prevalent HIV infection at baseline so as to examine incident HIV. We then further excluded prevalent HSV-2 cases or those without a follow-up HSV-2 test from our HIV cohort for the incident HSV-2 outcome.

Young women were seen annually from baseline until study completion or graduation from high school. Each annual study visit included an audio computer-assisted self-interview with the young woman and her parent/guardian and HIV and HSV-2 testing for young women who were not positive at the previous visit. Up to four assessments of the young woman and her parent/guardian were conducted from 2011 to 2015. An additional HIV and HSV-2 test was done around the time of the young woman's graduation for young women who missed a visit or graduated during the study. The median time from the previous visit to the additional test was 5 months (interquartile range: 4, 6). Young women aged over the study period with the highest age at the last survey being 23 years.

Information on the exposures of school attendance and dropout was collected directly from high school attendance registers. School attendance was defined as the average percentage of days attended in February, May, and August between follow-up visits. February, May, and

August were the months used to define dropout and attendance because these months were reported to be the most representative of normal attendance (because of lack of holidays/examinations) and data were collected for all young women during these months. Attendance was dichotomized as high ($\geq 80\%$ of school days on average between surveys) versus low ($< 80\%$ of school days) as the original cash transfer study provided the intervention based on this cut off [15]. The original intervention did not have an impact on incident infection or school attendance [15]. Young women who dropped out of school between surveys were defined as attending 0% of school days. The exposure of school dropout was defined as a report of dropout during any month between surveys. Dropout was time varying where young women were allowed to dropout during one year and return another. Incident HIV and incident HSV-2 infection were defined as detection of a new case of infection in a young woman previously testing negative. More information about diagnostic testing procedures is available elsewhere [15].

Statistical methods

To estimate the effect of school attendance on time to detection of both HIV and HSV-2, we used inverse probability of exposure weighted survival curves and a marginal structural Cox model [20,21]. We compared incidence of HIV and HSV-2 between those with high and low attendance using risk ratios, risk differences, and hazard ratios. We then compared incidence of HIV and HSV-2 between those who had dropped out in the previous period and those remaining in school using hazard ratios.

In these analyses, each young woman was followed from the date of her first negative HIV/HSV-2 test at enrollment until date of detection of infection, or date of censoring if she moved, was lost to follow-up, graduated, or reached the end of the study period. Young women who were lost to follow-up or died prior to completion of follow-up were censored as of their last valid test. We used the exposure and covariate values from the follow-up visit prior to outcome ascertainment to ensure that the outcome most likely occurred following exposure. When data were missing at a given visit, covariate data from the most recent observed value were carried forward. Most follow-up visits occurred between 6 to 18 months apart (median 12.3 months).

We first calculated crude risk of incident HIV and HSV-2 infection for young women with high versus low attendance using the extended Nelson–Aalen estimator of the survival function [22]. We then estimated the risk that all participants had high attendance and that all participants had low attendance throughout the study period using inverse probability of exposure weights (IPW) [23,24]. We used IPW to account for time-varying confounders and to produce weighted cumulative incidence curves that could be used to calculate risks,

risk differences, and risk ratios while accounting for time-varying shifts in the number of young women attending school over the study period [25].

We used these curves to calculate the risk ratio and risk difference of incident HIV and HSV-2 infection by school attendance at each year of follow-up (at 1.5, 2.5, and 3.5 years to capture later study visits and to account for the extra graduation visit). To capture the relationship between school attendance and the outcomes of interest, we additionally estimated a hazard ratio using a weighted Cox proportional hazards model. To estimate the association between school dropout and HIV/HSV-2, we estimated a weighted hazard ratio alone and did not produce weighted curves.

We used a directed acyclic graph to identify a minimally sufficient adjustment set of both time fixed and time-varying covariates. These covariates were used to calculate the weights for school attendance and dropout, which were used in both the weighted cumulative incidence curves and marginal structural Cox models. Confounders for incident HIV and HSV-2 included time-varying age at each visit, intervention arm, follow-up visit number, prior parental monitoring, prior partner age, prior pregnancy, prior attendance, and prior SES. Parental monitoring was constructed using a previously validated six item measure assessing perceived parental monitoring [26]. We also included prior HSV-2 status for the HIV outcome and prior HIV for the HSV-2 outcome. Additional covariates that were considered but not included in the final model included alcohol use, orphan status, school, children's depression inventory score [27,28], and revised children's manifest anxiety score [29]. We accounted for confounders in the minimally sufficient adjustment set using stabilized IPW [30]. Each young woman's IPW W_i at time t was the inverse of the probability of receiving the exposure that she did in fact receive, conditional on the vector L including confounders above,

$$W_{i,t} = \frac{P(A_{i,t}) = a}{P(A_{i,t}) = a | L_{i,t} = l}$$

The denominator of the IPWs were calculated using a pooled logistic regression model for each exposure including all confounders [21,31]. Numerator weights were estimated using a pooled logistic regression model for each exposure without covariates [30]. To adjust for potentially informative censoring, we multiplied our IPW by time-varying inverse probability of censoring weights. Censoring weights included the exposure as a predictor to account for loss to follow-up by exposure status. Pooled logistic regression models analogous to the previous IPW were used to estimate the numerator and denominator of censoring weights and weights were multiplied over time [20,21,32].

Confidence intervals for risk ratios and risk differences were calculated using the standard deviation from a bootstrap calculated from 200 full samples (with replacement) from the observed data [33,34]. Confidence intervals around the hazard ratios were created using the robust sandwich variance estimator [32]. The exact method was used to account for tied data and the proportional hazards assumption was evaluated by testing a product term between each exposure and time.

Results

Of the 2533 young women included in the parent study, 81 were excluded because they were HIV-positive at baseline and 124 were excluded because they did not have a follow-up test visit after baseline. Therefore, a total of 2328 young women were included in our analysis cohort for the outcome of incident HIV infection. From the 2328 young women in the HIV cohort, we further excluded 87 prevalent cases of HSV-2 at baseline and three who were missing HSV-2 status for a total of 2238 young women in the incident HSV-2 cohort.

Of the 2328 young women who were HIV-negative at baseline, 6.9% ($N=161$) ever dropped out between enrollment and the end of the study period. Of 6297 person visits over the study period, young women had low school attendance ($<80\%$ school days) at 5.4% ($N=341$) of visits. The median age at baseline was 15 years with 25.9% ($N=603$) reporting ever having had sex (Table 1). At baseline, 27.8% ($N=617$) of young women were orphans, 33.9% ($N=789$) had ever repeated a grade, 3.7% ($N=87$) had prevalent HSV-2 infection, and 8.3% ($N=190$) had ever been pregnant. Only 8.7% ($N=201$) ever used alcohol.

Over the study period, 107 incident HIV cases occurred among the 2328 young women in the HIV cohort, and 208 incident HSV-2 cases occurred among the 2238 young women in the HSV-2 cohort. Risk of both HIV and HSV-2 was higher for young women with low ($<80\%$) versus high attendance ($\geq 80\%$) in school (Figs. 1 and 2). After accounting for confounding, risk of incident HIV at the end of the study period was 7.6% for young women with high attendance versus 19.9% in young women with low attendance (Table 2). Risk of incident HSV-2 at the end of the study period was higher than HIV at 17.3% for young women with high attendance versus 38.5% in young women with low attendance, weighting for confounders.

The difference in cumulative incidence of HIV between those with low versus high attendance increased over time from 3.9% (95% CI: -0.9% , 8.7%) at 1.5 years to 12.3% (95% CI: 0.3% , 24.3%) at 3.5 years (Table 2). The difference in cumulative incidence of HSV-2 between

Table 1. Baseline characteristics of HIV-negative women aged 13 to 21 in Agincourt, South Africa from March 2011 to December 2012 ($N=2328$).

	<i>N</i> (%)	Median (IQR)
Young women's age at baseline (year)		15 (14,17)
Age 13–14	734 (31.5)	
Age 15–16	996 (42.8)	
Age 17–18	498 (21.4)	
Age 18–21	100 (4.3)	
Household wealth		
Low	587 (25.3)	
Middle to Low	620 (26.7)	
Middle	569 (24.5)	
High	548 (23.6)	
Intervention arm	1214 (52.2)	
Ever repeated a grade	789 (33.9)	
Partner 5 years or older	129 (5.5)	
Double or single orphan	617 (27.8)	
Maximum age difference with partner at baseline		0 (0,1)
Ever pregnant or had a child	190 (8.3)	
Ever had sex	603 (25.9)	
Unprotected sex in the last 3 months	181 (7.8)	
Given money or gifts in exchange for sex	82 (3.4)	
Prevalent HSV-2 infection	87 (3.7)	
Alcohol use	201 (8.7)	
Parental monitoring		10 (8,12)
Children's depression index score		2 (1,5)
Children's manifest anxiety score		4 (1,7)

Missing: age 0; SES 4; intervention arm 0; grade repetition 0; partner age difference 6; orphan 108; ever pregnant 28; ever sex 3; unprotected sex 19; lifetime partners 23; transactional sex 104; HSV-2 infection 3; alcohol use 4; parental monitoring 42; depression 115; anxiety 34. HSV-2, herpes simplex virus type 2; IQR, interquartile range; SES, socioeconomic status.

those with low versus high attendance was greater than HIV and also increased over time from 5.6% (95% CI: 0.4% , 10.8%) at 1.5 years to 21.2% (95% CI: 3.6% , 38.8%) at 3.5 years.

After accounting for confounders, young women with low attendance ($<80\%$) were more likely to develop HIV (HR: 2.97; 95% CI: 1.62, 5.45) and HSV-2 (HR: 2.47; 95% CI: 1.46, 4.17) over the entire follow-up period than young women with high attendance ($\geq 80\%$) (Table 3). Similarly, young women who dropped out of school had a higher weighted hazard of both HIV (HR: 3.25 95% CI: 1.67, 6.32) and HSV-2 (HR: 2.70; 95% CI: 1.59, 4.59) (Table 3). The effect of dropout was slightly greater than the effect of low versus high attendance in school on both incident HIV and HSV-2.

Discussion

Risk of HIV and HSV-2 increased over the study period and was higher in young women with low compared with high attendance in school. After accounting for confounding, the cumulative incidence of HIV at 3.5 years was 7.6% for young women with high attendance

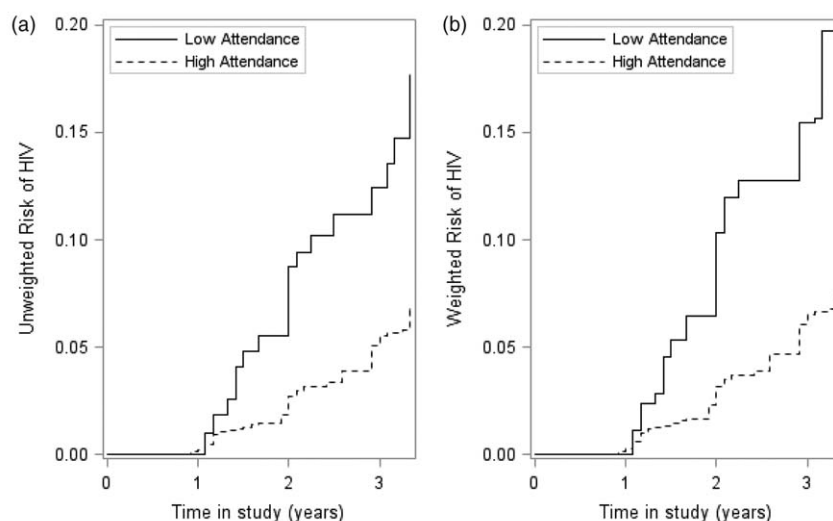


Fig. 1. Cumulative incidence curves for the association between school attendance and incident HIV infection for 2328 young HIV-negative women aged 13–23 participating in HPTN068 (2011–2015). (a) Unweighted; (b) weighted. Treatment weighted curves accounted for the following covariates: age, intervention arm, HSV-2, parental involvement, partner age, pregnancy, SES. HPTN, HIV prevention trials network; SES, socioeconomic status.

versus 19.9% in young women with low attendance. The cumulative incidence of HSV-2 at 3.5 years was higher than HIV at 17.3% for young women with high attendance versus 38.5% in young women with low attendance. The weighted hazard of HIV and HSV-2 was greater for young women who attended less school compared with more school, and who dropped out compared with those who stayed in school.

Our results are compatible with previous research showing a reduced risk of sexually transmitted infections

in young people attending school. Three studies, one in South Africa [10] and two in Zimbabwe [13,35] found that young women not attending school had three or more times the odds of HIV or HSV-2 infection compared with those who were attending school. More recently, a cash transfer intervention aimed at keeping girls in school in Malawi lowered prevalence of HSV-2 and HIV [36]. In the cash transfer study (HPTN 068) from which these data originate, the cash transfer intervention to keep girls in school did not have an effect on school attendance or HIV incidence, but overall

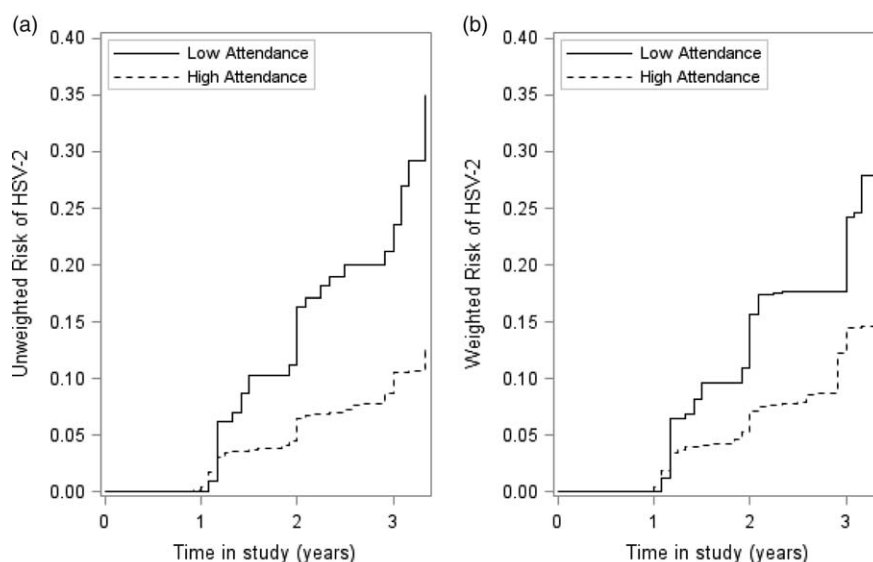


Fig. 2. Cumulative incidence curves for the association between school attendance and incident HSV-2 infection for 2238 young HIV and HSV-2-negative women aged 13–23 participating in HPTN068 (2011–2015). (a) Unweighted; (b) weighted. Treatment weighted curves accounted for the following covariates: age, intervention arm, HIV, parental involvement, partner age, pregnancy, SES. HPTN, HIV prevention trials network; HSV-2, herpes simplex virus type 2; SES, socioeconomic status.

Table 2. Weighted risk differences and risk ratios for the effect of school attendance on incident HIV and HSV-2 infection among young women aged 13–23, enrolled in HPTN 068 between 2011 and 2015 at different time intervals.

	Number infected	Risk (%)	Risk difference (%)	95% CI (%)	Risk ratio	95% CI
HIV						
1.5 years						
Low attendance (<80%)	6	5.3	3.9	(−0.9, 8.7)	3.71	(1.21, 11.38)
High attendance (≥80%)	27	1.4	0	0	1	1
2.5 years						
Low attendance (<80%)	14	12.8	8.9	(0.8, 17.0)	3.29	(1.61, 6.70)
High attendance (≥80%)	61	3.9	0	0	1	1
3.5 years						
Low attendance (<80%)	19	19.9	12.3	(0.3, 24.3)	2.62	(1.38, 4.97)
High attendance (≥80%)	88	7.6	0	0	1	1
HSV-2						
1.5 years						
Low attendance (<80%)	11	9.6	5.6	(0.4, 10.8)	2.40	(1.29, 4.47)
High attendance (≥80%)	76	4.0	0	0	1	1
2.5 years						
Low attendance (<80%)	22	17.6	9.7	(1.6, 17.8)	2.21	(1.34, 3.65)
High attendance (≥80%)	134	8.0	0	0	1	1
3.5 years						
Low attendance (<80%)	34	38.5	21.2	(3.6, 38.8)	2.23	(1.23, 4.04)
High attendance (≥80%)	174	17.3	0	0	1	1

Weighted curves conditioned on the following covariates: HIV: age, intervention arm, visit, HSV-2, parental involvement, partner age, pregnancy, previous attendance, SES; HSV-2: Age, intervention arm, HIV, parental involvement, visit, partner age, pregnancy, SES. CI, confidence interval; HPTN, HIV prevention trials network; HSV-2, herpes simplex virus type 2; SES, socioeconomic status.

attendance was high in the study population because of multiple factors [15]. Interventions to keep girls in school and to increase the frequency at which young women attend school could prevent HIV and HSV-2 but more research is needed to determine the best way to do so in varying contexts.

Prior studies have largely focused on the relationship between increased level of education or educational attainment and risk of HIV and HSV-2 among adults in sub-Saharan Africa [5–7,9,37–41]. Our results show that frequency of time spent in school among young women of high school age also has an effect on incident HIV and

HSV-2. The effect of dropout on both HIV and HSV-2 was larger than the effect of low versus high attendance. As dropout is an extreme of low attendance, the larger effect demonstrates that the less time a girl spends in school the further her risk is increased. Associations with frequency of school attendance in young women indicate that spending more time in school may impose time and sexual network constraints that reduce exposure to infection, and that these reductions may or may not be related to school performance. More research is needed to better understand the mechanism by which education reduces risk of HIV and HSV-2, but other analyses from our study in South Africa indicate that school attendance

Table 3. Hazard ratios for the effect of school attendance and school dropout on incident HIV and HSV-2 infection among young women aged 13–23, enrolled in HPTN 068 from 2011–2015.

	Unweighted				Weighted ^a			
	No. of events	Person-months of follow-up	HR	95% CI	No. of events	Person-months of follow-up	HR	95% CI
HIV								
Total	107	164156			107	148917		
Low school attendance (<80%)	19	10930	2.82	1.70, 4.66	16	7164	2.97	1.62, 5.45
High school attendance (≥80%)	88	153226	1	1	91	141753	1	1
Yes dropout	16	7604	3.24	1.87, 5.63	16	6167	3.25	1.67, 6.32
No dropout	91	156349	1	1	95	142373	1	1
HSV-2								
Total	208	154698			229	140491		
Low school attendance (<80%)	34	9329	3.08	2.16, 4.40	30	7589	2.47	1.46, 4.17
High school attendance (≥80%)	174	145369	1	1	199	132902	1	1
Yes Dropout	183	6308	3.22	2.15, 4.83	23	5333	2.70	1.59, 4.59
No Dropout	25	148228	1	1	194	135140	1	1

CI, robust confidence intervals; HPTN, HIV prevention trials network; HR, hazard ratio; HSV-2, herpes simplex virus type 2.

^aWeighted models conditioned on the following covariates: HIV: age, intervention arm, HSV-2 status, parental involvement, partner age, previous attendance, visit, pregnancy, SES. HSV-2: age, intervention arm, HIV status, parental involvement, partner age, pregnancy, visit, SES.

affects partner selection including both partner age difference and number of sexual partners [Stoner *et al.*, forthcoming].

In the HPTN 068 study, HIV and HSV-2 were measured at each follow-up visit roughly a year apart. Although the specific date of infection is unknown, we are certain that the infection occurred within the year between visits. It is possible that infection occurred while school attendance was higher or lower within this interval or before dropout. This could have led to issues of reverse causality. However, HIV and HSV-2 testing do not occur regularly enough in the community that it would be common for young women to know their status and then dropout of school. It is more likely for young women to dropout and then have an infection that is identified through participation in the study or later testing. Additionally, a sensitivity analysis using time of infection two visits later (i.e. attendance between baseline and visit 1 with time of infection at visit 2) yielded a similar association between attendance/dropout and risk of infection. Last, although acute infection could lead to sick days (low attendance), a greater number of studies have explored schooling as an exposure leading to infection rather than vice versa [10–13].

All young women were in school at study enrollment, which may make our results less generalizable to young women who were not enrolled but later returned to school, or those who were not accessing school for any other reasons. In addition, there may also have been a Hawthorne effect of the trial where young women were less likely to dropout because of trial participation [42]. Second, information on sexual behaviors and partner characteristics were self-reported in the study and may have been misreported. To minimize reporting bias, interviews were conducted in a private setting using audio computer-assisted self-interview [43]. Third, our analysis assumes no unmeasured confounding, which is a strong assumption that may not be valid. However, we did include key potential confounders based on the literature, including pregnancy, age, and SES. We also considered other confounders including school, mental health status, alcohol use, and orphanhood that were not included in the final set of confounders and did not substantially alter estimates. Sexual or physical violence were not considered because detailed information was only collected for young women who were sexually active.

Interventions to increase frequency of school attendance and prevent dropout should be promoted to reduce risk of infection. The reduced risk of infection may be because of the time and network constraints imposed by being in a school environment in addition to gains in formal education, but more research is needed to understand the mechanisms for the relationship. Additionally, more research is needed to understand the best way to prevent dropout and low attendance in different settings. Our

study adds to the literature by using longitudinal data to examine incident infection in relation to the frequency of time that young women spend in a school environment instead of educational attainment in women who have completed their education. Staying in school and attending more school days are associated with a reduced risk of incident HIV and HSV-2.

Acknowledgements

We thank the HPTN 068 study team and all trial participants. The MRC/Wits Rural Public Health and Health Transitions Research Unit and Agincourt Health and Socio-Demographic Surveillance System have been supported by the University of the Witwatersrand, the Medical Research Council, South Africa, and the Wellcome Trust, UK (grants 058893/Z/99/A; 069683/Z/02/Z; 085477/Z/08/Z; 085477/B/08/Z). Additionally, we would like to thank William C. Miller for his substantial contributions to the manuscript conception. All authors contributed to drafting and editing the manuscript, interpretation of the data, approved the final version and are accountable for all aspects of the work. A.P., J.K.E., A.E.A., and C.T.H. contributed to the conception and design of the analysis. The remaining authors were involved in data acquisition, data collection, study management, and design of the original parent study. Additionally, J.W., J.K.E., and J.P.H. contributed to the analysis of data.

The study was funded by T32 5T32AI007001 from the National Institutes of Health (NIH) and by Award Numbers UM1 AI068619 (HPTN Leadership and Operations Center), UM1AI068617 (HPTN Statistical and Data Management Center), and UM1AI068613 (HPTN Laboratory Center) from the National Institute of Allergy and Infectious Diseases, the National Institute of Mental Health and the National Institute on Drug Abuse of the National Institutes of Health. This work was also supported by NIMH R01 (R01MH087118) and the Carolina Population Center and its NIH Center grant (P2C HD050924). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of interest

There are no conflicts of interest.

Accepted to International AIDS Society (IAS) conference 2017. Paris, France: July 23–26, 2017.

References

1. Gómez-Olivé FX, Angotti N, Houle B, Klipstein-Grobusch K, Kabudula C, Menken J, *et al.* **Prevalence of HIV among those 15 and older in rural South Africa.** *AIDS Care* 2013; **25**:1122–1128.

2. Jewkes R, Wood K, Duvvury N. I woke up after I joined Stepping Stones': meanings of an HIV behavioural intervention in rural South African young people's lives. *Health Educ Res* 2010; **25**:1074–1084.
3. de Walque D. How does the impact of an HIV/AIDS information campaign vary with educational attainment? Evidence from rural Uganda. *J Dev Econ* 2007; **84**:686–714.
4. Hargreaves JR, Howe LD. Changes in HIV prevalence among differently educated groups in Tanzania between 2003 and 2007. *AIDS* 2010; **24**:755–761.
5. Michelo C, Sandøy IF, Fylkesnes K. Marked HIV prevalence declines in higher educated young people: evidence from population-based surveys (1995–2003) in Zambia. *AIDS* 2006; **20**:1031–1038.
6. De Neve JW, Fink G, Subramanian SV, Moyo S, Bor J. Length of secondary schooling and risk of HIV infection in Botswana: evidence from a natural experiment. *Lancet Glob Health* 2015; **3**:e470–e477.
7. Bärnighausen T, Hosegood V, Timaeus IM, Newell ML. The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa. *AIDS* 2007; **21** (Suppl 7):S29–S38.
8. Fleischman J, Peck K. Addressing HIV Risk in Adolescent Girls and Young Women.; 2015. http://csis.org/files/publication/150410_Fleischman_HIVAdolescentGirls_Web.pdf. [Accessed 20 December 2016].
9. Jukes M, Simmons S, Bundy D. Education and vulnerability: the role of schools in protecting young women and girls from HIV in southern Africa. *AIDS* 2008; **22** (Suppl 4):S41–S56.
10. Pettifor AE, Levandowski BA, MacPhail C, Padian NS, Cohen MS, Rees HV. Keep them in school: The importance of education as a protective factor against HIV infection among young South African women. *Int J Epidemiol* 2008; **37**:1266–1273.
11. Hargreaves JR, Morison LA, Kim JC, Bonell CP, Porter JD, Watts C, et al. The association between school attendance, HIV infection and sexual behaviour among young people in rural South Africa. *J Epidemiol Community Health* 2008; **62**:113–119.
12. Baird S, Chirwa E, McIntosh C, Ozer B. The short-term impacts of a schooling conditional cash transfer program on the sexual behavior of young women. *Health Econ* 2010; **19**:55–68.
13. Gavin L, Galavotti C, Dube H, McNaghten AD, Murwirwa M, Khan R, St Louis M. Factors associated with HIV infection in adolescent females in Zimbabwe. *J Adolesc Health* 2006; **39**:596e11–596e18.
14. Stroeken K, Remes P, De Koker P, Michielsens K, Van Vossle A, Temmerman M. HIV among out-of-school youth in Eastern and Southern Africa: a review. *AIDS Care* 2012; **24**:186–194.
15. Pettifor A, MacPhail C, Hughes JP, Selin A, Wang J, Gómez-Olivé FX, et al. The effect of a conditional cash transfer on HIV incidence in young women in rural South Africa (HPTN 068): a phase 3, randomised controlled trial. *Lancet Glob Health* 2016; **4**:e978–e988.
16. Pettifor A, MacPhail C, Selin A, Gomez-Olive FX, Rosenberg M, Wagner RG, et al. HPTN 068 protocol team. HPTN 068: a randomized control trial of a conditional cash transfer to reduce HIV infection in young women in South Africa: study design and baseline results. *AIDS Behav* 2016; **20**:1863–1882.
17. Collinson MA, Tollman SM, Kahn K. Migration, settlement change and health in post-apartheid South Africa: triangulating health and demographic surveillance with national census data. *Scand J Public Health Suppl* 2007; **69**:77–84.
18. Kahn K, Tollman SM, Collinson MA, Clark SJ, Twine R, Clark BD, et al. Research into health, population and social transitions in rural South Africa: Data and methods of the Agincourt Health and Demographic Surveillance System1. *Scand J Public Health* 2007; **35**:8–20.
19. Sartorius K, Sartorius BK, Collinson MA, Tollman SM. The dynamics of household dissolution and change in socio-economic position: a survival model in a rural South Africa. *Dev South Afr* 2014; **31**:775–795.
20. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 2000; **11**:561–570.
21. Cole SR, Hernán MA, Robins JM, Anastos K, Chmiel J, Detels R, et al. Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. *Am J Epidemiol* 2003; **158**:687–694.
22. Cole SR, Hudgens MG. Survival analysis in infectious disease research: describing events in time. *AIDS* 2010; **24**:2423–2431.
23. Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med* 2005; **24**:3089–3110.
24. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed* 2004; **75**:45–49.
25. Xu S, Shetterly S, Powers D, Raebel MA, Tsai TT, Ho PM, Magid D. Extension of Kaplan-Meier methods in observational studies with time-varying treatment. *Value Health* 2012; **15**:167–174.
26. Li X, Stanton B, Feigelman S. Impact of perceived parental monitoring on adolescent risk behavior over 4 years. *J Adolesc Health* 2000; **27**:49–56.
27. Cluver L, Gardner F, Operario D. Psychological distress amongst AIDS-orphaned children in urban South Africa. *J Child Psychol Psychiatry* 2007; **48**:755–763.
28. Kovacs M. The children's depression inventory (CDI). *Psychopharmacol Bull* 1985; **21**:995–998.
29. Boyes ME, Cluver LD. Performance of the revised children's manifest anxiety scale in a sample of children and adolescents from poor urban communities in Cape Town. *Eur J Psychol Assess* 2013; **29**:113–120.
30. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008; **168**:656–664.
31. D'Agostino RB, Lee ML, Belanger AJ, Cupples LA, Anderson K, Kannel WB. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Stat Med* 1990; **9**:1501–1515.
32. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**:550–560.
33. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat Med* 2000; **19**:1141–1164.
34. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci* 1986; **1**:54–75.
35. Birdthistle I, Floyd S, Nyagadza A, Mudziwapasi N, Gregson S, Glynn JR. Is education the link between orphanhood and HIV/HSV-2 risk among female adolescents in urban Zimbabwe? *Soc Sci Med* 2009; **68**:1810–1818.
36. Baird SJ, Garfein RS, McIntosh CT, Özler B. Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: A cluster randomised trial. *Lancet* 2012; **379**:1320–1329.
37. Hargreaves JR, Bonell CP, Boler T, Boccia D, Birdthistle I, Fletcher A, et al. Systematic review exploring time trends in the association between educational attainment and risk of HIV infection in sub-Saharan Africa. *AIDS* 2008; **22**:403–414.
38. de Walque D, Nakiyingi-Miiró JS, Busingye J, Whitworth JA. Changing association between schooling levels and HIV-1 infection over 11 years in a rural population cohort in south-west Uganda. *Trop Med Int Health* 2005; **10**:993–1001.
39. Brent RJ. A cost-benefit analysis of female primary education as a means of reducing HIV/AIDS in Tanzania. *Appl Econ* 2009; **41**:1731–1743.
40. Hargreaves JR, Glynn JR. Educational attainment and HIV-1 infection in developing countries: a systematic review. *Trop Med Int Health* 2002; **7**:489–498.
41. Glynn JR, Caraël M, Buvé A, Anagonou S, Zekeng L, Kahindo M, Musonda R, Study Group on Heterogeneity of HIV Epidemics in African Cities. Does increased general schooling protect against HIV infection? A study in four African cities. *Trop Med Int Health* 2004; **9**:4–14.
42. Rosenberg M, Pettifor A, Twine R, Hughes J, Gómez-Olivé FX, Wagner RG, et al. Selection and Hawthorne effects in an HIV prevention trial among young South African women. *International Aids Society (IAS)*. Durban, South Africa; 2016.
43. Morrison-Beedy D, Carey MP, Tu X. Accuracy of audio computer-assisted self-interviewing (ACASI) and self-administered questionnaires for the assessment of sexual behavior. *AIDS Behav* 2006; **10**:541–552.